The effects of network structure and geography on the spread of epidemics in finite populations.

Kieran Gourley

February 14, 2008

Abstract

In this paper we create a new model of disease transmission that combines a branching process from epidemiology with a geographic network structure. We present two versions of this geographic network model, the first is a simplified network model based on a 2-dimensional regular lattice. While the second is a more complex network structure with long distance connections. We present a series of numerical results from our two models and find that for the 2D lattice that the proportion of the total population infected by a disease depends strongly on the total population size. In the long distance model there is no such dependence and we find that two infection states exist; either a large proportion of the population is infected or only a small proportion of the population is infected.

Introduction

Turn on the news today and you will hear stories about computer viruses, bird flu, hit records, fashion trends and disease outbreaks. What do they have in common, why are they important and can we understand them? We live in a connected world where the effects of globalisation are part of our everyday lives. To understand our world we need to understand the networks that surround us. In this paper we look at networks and specifically the role they play in transmitting information (objects) through a population. These objects can take multiple forms depending on what we are studying; from a computer virus containing malicious code to the molecular sequence of a new viral strain or the compressed audio of an mp3 record. The objects all share the property that they can be replicated and transmitted between hosts that are connected by some type of network. In this paper we will look at the geographical structure of complex networks and the way in which geography affects the spread of ideas, diseases and objects.

The outline of the paper is as follows, in the first part of the Background section we look at some of the existing work that has been done in epidemiology. Then in the second part of the Background section we outline some of the research being done in the field of networks and network theory. The results section is

split into two parts, the first part introduces our network based model of information transmission and shows how a branching process works on a simple 2D lattice. The second part of the results section introduces a more complex geographic network model with long distance connections. We discuss the numerical results from these two models and the final section discusses the findings and conclusions of our work.

Background

Research in the field of epidemiology has traditionally been concerned with the spread of a disease within a population. However this approach can be equally well applied to the spread of ideas, information or objects within the same population. In modelling the spread of a disease it is standard practice to divide the population (agents) into different categories depending on their response to the disease. There are then four different classes of agents where each class corresponds to the infection status of the agents.

1. S - Susceptible to infection
2. E - Exposed to infection but still latent
3. I - Infected, carrying the disease and infectious to other agents
4. R - Removed, the agent has either recovered or died from the disease

This is known as the SEIRS model of infection. Where the different letters describe the movement of an agent between classes. In some diseases there is no latent period in which case the model reduces to a simpler SIR model, whilst in other cases there is no recovery and we have a SEI model. The amount of time spent in each stage varies for different diseases and can affect the dynamics of the transmission.

Although there can be variations in the specifics of the model a common feature to most epidemiological models is a branching process of infection. Where one agent within the population infects R_0 new agents and each of these agents in turn infects R_0 more agents. This process is shown schematically in Fig.1. Here R_0 is the basic reproductive rate of the infection. Or more formally, R_0 is the average number of secondary infections produced by one infected individual. From this we can then see that an infection can only maintain itself within a population when $R_0 > 1$.

Within the branching process model, it is traditionally assumed that R_0 (the mean number of infections caused by an infected individual in a fully susceptible population) is derived from a normal distribution. However work done by May et al. (1987) on the spread of the HIV virus found that variations in the distribution of R_0 can lead to massive changes in the number of individuals infected. The work by May et al. showed that when R_0 was normally distributed the total number of individuals who were infected was relatively small. However, when R_0 was changed to a more realistic fat tailed distribution, then the number of individuals infected with HIV increased dramatically.

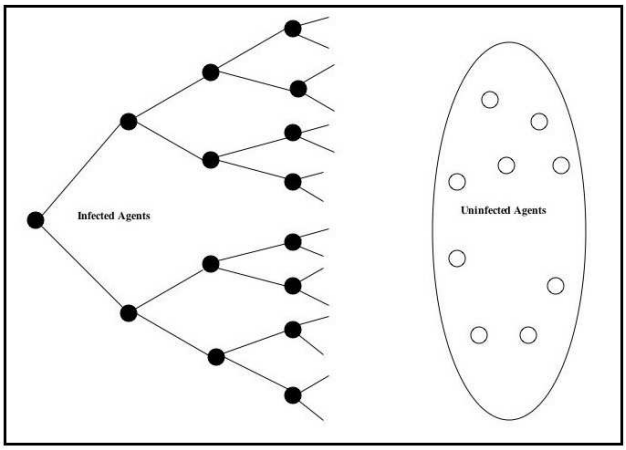

Figure 1: A schematic of the branching process in which R_0 is constant. In this instance $R_0 = 2$, though in general R_0 can be given a value from any distribution.

Lloyd-Smith model

The work of May et al (1987) was extended in the 2005 paper by Lloyd-Smith et al. In this paper the authors found that by using a non-gaussian distribution of R_0, that superspreading is a normal feature of disease spread. Where an SSE is defined to be in the 99^{th} percentile of secondary infection cases. In the Lloyd-Smith (LS) model the authors introduced the parameter ν, an individual reproductive number, as a random variable representing the expected number of secondary cases caused by a particular infected agent. Continuous probability distributions which contained all the information on the variation in infectious histories of individuals, including differences in the host, pathogen and environment were then used to assign values to ν. By selecting three different continuous distributions they created three different offspring distributions (the Poisson, geometric and negative binomial) and used them to analyse contact tracing data from eight directly transmitted diseases.

They found the Poisson offspring distribution, where there is is only a small individual variation in the number of people infected, was almost always strongly rejected. In contrast, the geometric and negative binomial offspring distributions, which are drawn from exponential and gamma distributions respectively, contain a large degree of individual variation between agents. Because of this individual variation the geometric and binomial offspring distributions had support from a number of empirical data sets. The authors were able to conclude from these results that there was no real meaningful "average" number of people who could be infected by a contagious individual. Instead they found that a lot of people do not infect anyone and certain individuals can infect unusually large numbers of secondary cases. Lloyd-Smith et al (2005) then developed their mathematical model using the negative binomial offspring distribution to predict the disease dynamics arising from these superspreaders, and showed that when individual variation

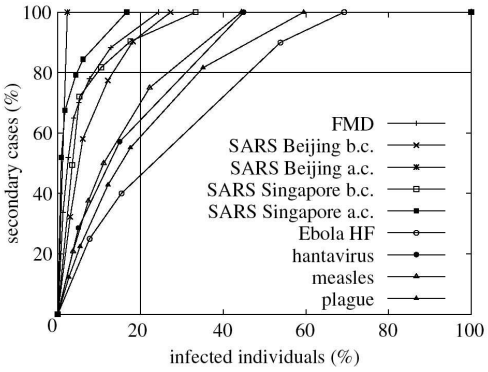

Figure 2: Cumulative percentage of secondary cases caused by the most infectious individuals for various disease outbreaks. The vertical axis represents the percent of secondary cases and the horizontal axis the percent of infected individuals. Abbreviations: bc=before control, ac=after control.

is large, an outbreak will either quickly die out or it will explode within the population. Using these models, they were then able to explore implications for outbreak control, showing that individual-specific control measures outperform population-wide measures.

Superspreading event based model (SSE)

James et al (2007) built on the work of Lloyd-Smith et al (2005) and introduced a new model of disease transmission known as the cumulative process (CP) model. Where Lloyd-Smith et al (2005) treated superspreaading as a demographic phenomenon where each individual had an infectivity allocated at random from a gamma distribution. In the CP model, James et al (2007) kept a normal distribution for R_0 and introduced superspreading events (SSE) that occur as a stochastic consequence of the complex network of interactions made by individuals. These SSEs could be created by any agent (as every agent had a finite probability of infecting a large number of other agents).

In the CP model infections can occur in one of two different ways each occurring with a specific intensity or frequency. The first of these processes is an infection from the standard Poisson branching process described in Fig.1 which occur with an intensity r. The second process of infection comes from distinct SSEs which occur with an intensity given by ρ. For SSEs in the CP model the number of infections is also Poisson distributed with a mean given by λ. Infections in the CP model last for a deterministic time $t_1 = 1$ and all infectious individuals are independent. From these conditions the authors were able to describe the offspring using a cumulative process with the following probability generating function

$$G(s) = \exp\{r(s-1) + \rho(\exp(\lambda(s-1)) - 1)\}. \tag{1}$$

With a mean and variance given by the following equations;

$$\mathrm{E(Z)} = (r + \rho\lambda), \qquad \mathrm{Var(Z)} = (r + \rho\lambda + \rho\lambda^2). \tag{2}$$

In the CP model the importance of SSEs is given by the variable γ which the authors define to be;

$$\gamma = \frac{\rho\lambda}{R_0}(1 - e^{-\lambda}(1 + \lambda)) \tag{3}$$

Using this dual infection based model James et al. were then able to compare the numerical results from the model against contact tracing data from real world disease oubreaks. They found that for certain classes of disease outbreaks that satisfy the 80/20 rule (namely 80% of the infections are caused by 20% of the individuals), there model was in good agreement with the empirical results. For disease outbreaks that did not satisfy the 80/20 rule a simpler model would be more appropriate to describe the transmission dynamics. A summary of the results from James et al. is shown in Fig.2 and we observe that SARS and FMD would be good candidate disease outbreaks to use the CP model on.

As far as control strategies go, the main aim has been to reduce R_0. However, we see with the advent of superspreading and SSEs this may not in fact be the most effective method of controlling a virus. James et al (2007) shows that non-SSE infections can have relatively little impact on the spread of a disease hence they believe there is an opportunity to develop control strategies aimed at reducing (a) the frequency and (b) the severity. This could be achieved by reducing the number of times agents meet and by breaking large events into smaller ones respectively.

Loyd-Smith et al (2005) and James et al (2007) both approach superspreading from different perspectives yet they seem to agree on a number of epidemiological outcomes one being that with increased superspreading, both see the virus becoming extinct more often but also outbreaks as potentially much more severe than a traditional R_0 analysis would give. As James et al. (2005) note, reality probably lies somewhere in between these two approaches and the need to marry the two is a vital next step.

James et al (2007) have shown that the numerical results from LS model can be explained by the existence of a few relatively rare super-spreading events in a population with an otherwise normally distributed series of contacts. However, despite the success of both the LS and CP models in reproducing empirical results from a range of disease outbreaks (see Fig.3), the models are still approximations that can be improved on as they assume infinite populations and ignore the structure of the infection/contact networks. Indeed James et al (2007) identify their models limitations in their conclusion:

> "...the roles of a finite population of susceptibles, and any explicit spatial structure, require further (possibly disease-specific) modeling outside the scope of this communication."

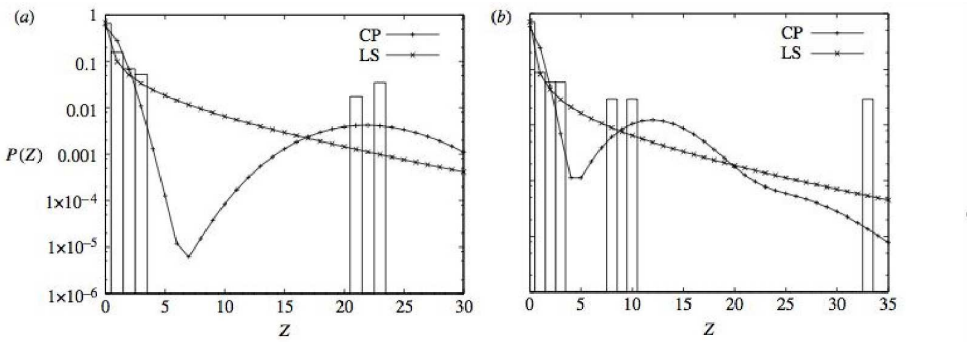

Figure 3: Offspring distributions for the SARS outbreak before control: (a) Singapore, 2003; (b) Beijing, 2003. $P(Z)$ is the probability of an infected individual causing Z secondary infections during their infectious lifetime. The empirical data is shown here as bars, the numerical results from the CP and LS models are shown as points.

Networks

Small-world networks

Over the last five years network theory has enjoyed a high profile in many different disciplines. Network theory looks at the structure of connections between nodes. In this section we investigate the structure of two major network models; the 'small world' research by Watts et al. (1998) and the scale free network models by Barabasi et al. (1999).

In basic network theory there are two types of objects, nodes and edges. The nodes are connected to each other by the edges and the collection of nodes and edges defines the network. For example a social network would make the nodes equal to people and the edges would then represent friendship ties, though this can change depending on the system being modelled. Before the paper by Watts et al. most of the networks being studied were either ordered (e.g. 2D lattice with nearest neighbour connections) or disordered (e.g. Erdos-Renyi (ER) random graph). Watts introduced a simple network model that occupied the topological space between order and disorder. The model they introduced is known as the small world network model.

To construct a small world network Watts first created a ring lattice with n vertices and k edges per vertex. This ordered network with $k = 4$ is simply a 1D network with nearest and next-nearest neighbour connections. From this ordered network Watts then introduced disorder through a random rewiring procedure such that each edge in the network has a probability p of being randomly rewired. These random links act as shortcuts across the network connecting two nodes that would otherwise be far apart. Thus, as p approaches unity the graph moves towards a topology described by the ER random graph. This small world network and rewiring process is shown schematically in Fig.4.

In order to understand the topology of the small-world network Watts defined two metrics (1) the

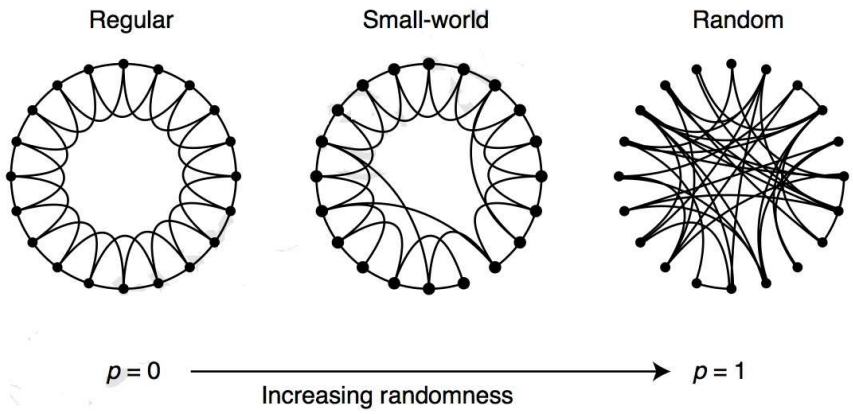

Figure 4: Small-world graph schematic showing the random rewiring process as p increases. Here $n = 20$ and $k = 4$.

characteristic path length $L(p)$ and (2) the clustering coefficient $C(p)$. Where $L(p)$ measures the standard separation between two vertices and $C(p)$ measures the clustering within a typical neighbourhood. When $p = 0$ then there is a high clustering coefficient ($C(p)$), but the path length between two random vertices ($L(p)$) is large. Whilst when $p = 1$, $L(p)$ is small meaning that information can travel quickly across the network, but there is very little in the way of a 'community' structure as $C(p)$ is low. These two limiting cases seem to reveal that clustering is always associated with long path lengths, whilst a short path length requires disorder. However Watts found that with the addition of a few randomly rewired links (i.e. small p) that a network could exist with both a high $C(p)$ *and* a low $L(p)$. The long distance links effectively create short-cuts across the network that reduce the characteristic path length, but because there are only a few of them the overall clustering structure remains largely intact.

Mathematically the small world network is expressed as the diameter of the network growing with the logarithm of the number of nodes (rather than proportional to the number of nodes, as in the case for a lattice). These type of behaviour is very important for the understanding of disease transmission, since the addition of only a few random links can greatly reduce the characteristic path length which in turn results in a large increase in the rate of transmission.

Scale-free networks

The second major model of network growth and structure is the Barabasi-Albert (BA) model of scale-free networks. The BA model was initially created to describe the seemingly unusual network typology of the world wide web. On the world wide web the webpages are the nodes and the html links between them are directed edges. By studying the html links between webpages, Barabasi found the unusual feature that the number of links into a webpage did not follow a normal distribution. Instead most webpages had very few

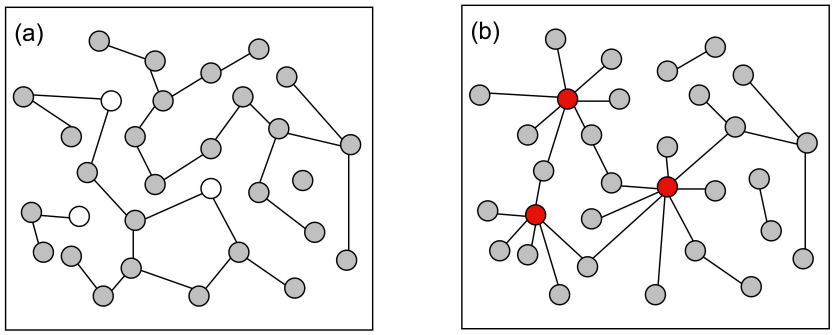

Figure 5: (a) random graph where each node has a probability p of being connected to any other node (b) scale-free graph generated from the BA model where links are biased towards nodes with a high number of links.

inward links whilst some webpages had literally millions. The precise distribution was found to follow a power-law distribution of the form;

$$P(x) = Cx^{-\alpha}, \tag{4}$$

where C is a constant, x is the number of links and α represents the slope of the line when plotted on a log-log graph.

The BA model attempts to recreate this network typology by using a simple set of underlying rules for connections between nodes. The BA model begins with an initial population of m_0 nodes and new nodes are added to the network one at a time. Each time a new node is added it is connected to an existing node with a probability that is proportional to the number of links the existing node already has. Formally this can be represented by the following expression;

$$p_i = \frac{k_i}{\sum_j k_j}, \tag{5}$$

where k is the number of connections that each node has, j is the total number of nodes in the network and p_i is the probability of linking the new node to node i. From this equation we can see that the BA model creates this power-law distribution by biasing the addition of links towards nodes that already have a large number of existing edges, i.e. the more links a node has, the more likely they are to get another link. This process can be thought of as a 'rich get richer' model of link formation. A schematic of a scale-free network can be seen in Fig.5(b). It is characterised by the existence of central hubs which contain a large number of connections. This topology is in stark contrast to that of a typical ER graph (see Fig.5(a)) which does not see the emergence of hubs.

The degree distribution resulting from the simple implementation of the BA model (i.e. nodes cannot

have links removed etc.) is of the form;

$$P(k) \sim k^{-3} \tag{6}$$

In addition to describing the degree distribution of the world wide web, the BA model also does a good job of describing many real-world networks including; the collaborative network of Hollywood actors (actors are nodes and edges form when two actors star in the same movie); the power-grid of the western United States (power stations are nodes and links are the transmission lines); and peer-reviewed scientific literature (nodes are publications and citations are the links). Networks that exhibit a power-law distribution of links are also very interesting from a disease transmission perspective since if a node with a high number of links gets infected then the disease can quickly spread to the rest of the population.

Both the scale-free and small world networks have been very successful in describing the structure of real world networks. Until recently only a small amount of research has been undertaken looking at the functions of networks i.e. how information/objects spread on these networks. Recent work by Newman et al (2004) has looked at this phenomenon when investigating the spread of viruses on networks. There has also been very little research investigating the effect that geography has on the spread of viruses on networks. Geography plays an important role in the spread of viruses. This was noted by May and Anderson (1987) where they identified truck drivers as the key transmitters of the HIV virus in Africa. These agents have a large number of contacts (high R_0), and they also travelled large distances. The two factors, connections and geography, came together to produce the HIV epidemic we see today. In this paper we create a model that examines the effects of network structure and geography on the spread of viruses/information in finite populations.

Results - geographic model

Simple 2D lattice network

To understand the effects of geography on the transmission of diseases we first constructed a simple network model with a spatial distribution of nodes. This network takes the form of a bounded two dimensional (2D) lattice with nearest neighbour connections (i.e. left/right and up/down). There are n agents/nodes in the network and these agents are arranged in a symmetrical $\sqrt{n}$ x $\sqrt{n}$ array. Away from the boundaries of the population the agents connections are homogeneously distributed with $k = 4$. We chose to use boundary conditions in the formation of the network as most real world problems have some form of boundary. This network is similar in nature to the 1D small world with network $p = 0$.

Once the network was generated the second step of the model involved running a branching process on the 2D ordered network. Here we used a simplification of the SEIR epidemiological model of infection and

set up three categories or states for the agents; S - susceptible, I - infected and R-removed. This would correspond to a disease where there is no long period of incubation corresponding to the E state. In the SIR model we allowed the agents to be in the I state for one time-step and then they moved into the R state where they were infected but could no longer infect any other agents.

From the population of n agents one agent i was randomly chosen and infected with the virus, this agent is known as the seed agent. At time-step $t = 1$, agent i has k_i connections and there is a probability p of agent i transmitting the disease to their nearest neighbour connections. If an infection occurs then the agent moves from the state S to state I and any agent previously in state I now move to the recovered/removed state R. Then at time-step $t = t + 1$ the infection process is repeated for any agent in state I. This process was run over multiple time-steps/generations and the total number of agents infected at each time-step was recorded. The simulation was run until either (a) the entire population was in state R, or (b) there were no longer any agents in state I (this corresponds to the disease dying out). We denote the proportion of the population that have been infected by the variable n_β.

We ran the simulation for varying population sizes to examine the effects of finite population on the spread of disease and we also varied p to understand how increased transmission rates altered the results. The results for the simple 2D nearest neighbour network with varying n are shown in Fig.6. The results are displayed on a log-log graph in the form of a histogram with the x-axis representing the proportion of the total agents infected n_β and the y-axis showing the proportion of total runs with this number of infections i.e. $P(n_\beta)$. For all values of n it is clear that the majority of runs infect only a small fraction of the population. As n_β increases, we initially observe a corresponding decrease in $P(n_\beta)$ irrespective of the size of the population. For large values of $n > 200$, this decreases in $P(n_\beta)$ continues in an almost linear fashion on the log-log graph until they reach a critical point where the decay is no longer linear. The size of this critical point decreases as n increases. For small values of $n < 200$ we start to observe finite size effects for large values of n_β. This is characterised by the flattening out or increasing of the curve for $n_\beta \sim 1$. This corresponds to an disease where the majority of the agents in the population are infected. For small population sizes this occurs on a disproportionately large scale when compared to large population groups.

These results were then plotted as a cumulative distribution and the 80^{th} percentile was calculated from this. The 80^{th} percentile is denoted by the variable N^{80} and can be thought of as the minimum number of people infected by the top 20 percent of all runs, i.e. if we were to infect the population 100 times then on average for 20 runs we would have at least N^{80} infections. These results are shown in Fig.7 on a log-log plot for the various values of n. From this we can see that as the population size increases the size of N^{80} increases dramatically. For $n = 9$ the top 20% of the runs infect at least 0.8 of the population, whilst for $n = 40,000$ the top 20% of the runs infect only $\sim 0.01\%$ of the population. From this we can conclude that on a 2D lattice population size plays a significant role in determining the probability of a disease outbreak

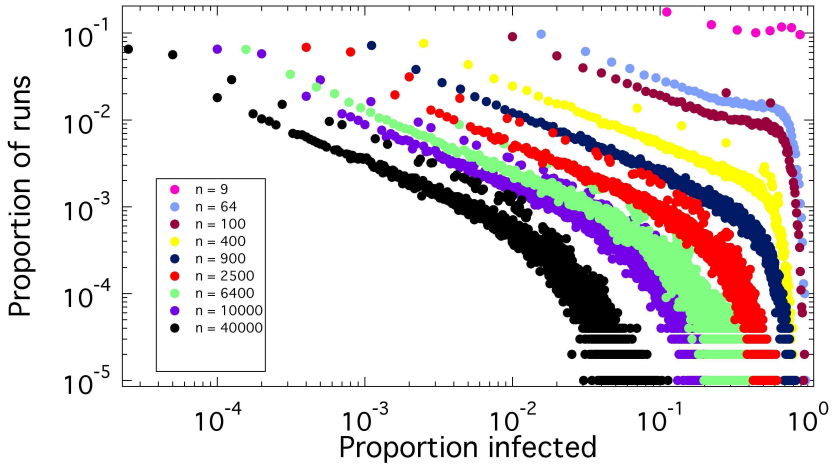

Figure 6: Numerical results for the transmission of a disease on a 2D network. The simulation was run 100,000 times for each population size n with nearest neighbour connections and $p = 0.5$. The x-axis records the total number of agents infected as a proportion and the y-axis records the proportion of runs with this number of infections.

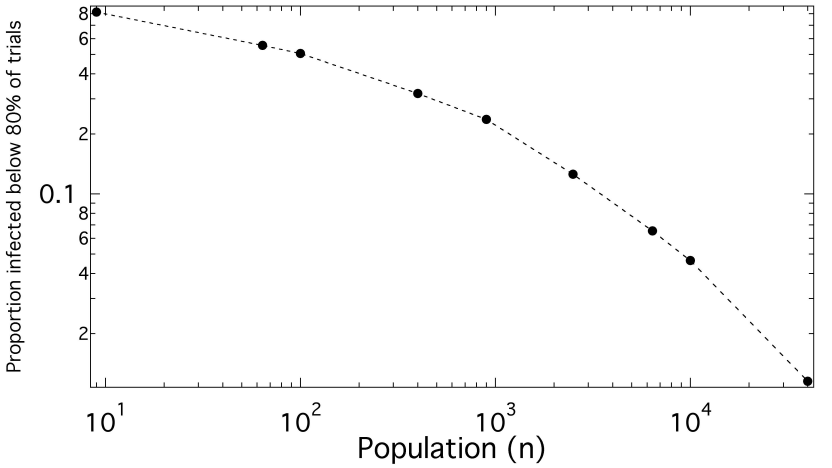

Figure 7: Graph showing N^{80} or the proportion of the population infected by the top 20% of the runs for a given population size. The simulation was run 100,000 times for different values of n and each circle represents the 80 percentile calculated from each set of data.

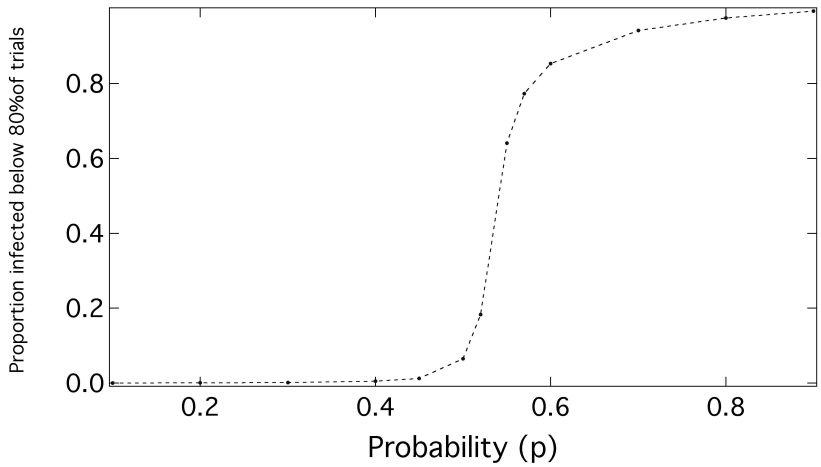

Figure 8: Graph showing N^{80} or the proportion of the population infected by the top 20% of the runs for a given value of p. The simulation was run 100,000 times for different values of n and each circle represents the 80 percentile calculated from each set of data.

infecting a large proportion of the population.

A similar set of numerical experiments was conducted for the same 2D network, but this time n was kept fixed and p the probability of transmission was varied. For each value of p 100,000 simulations were run and the total number of infections was recorded for each run. From this we can then calculate N^{80} and the results from this are displayed in Fig.8 for a population of size $n = 6400$. From these results we observe a distinct transition around $p_{\text{critical}} \sim 0.5$. Below this critical value N^{80} is small, which means that even for the most infectious realisations of the system, only a small proportion of the population is actually infected. Above the critical value things start to change dramatically as N^{80} increases rapidly. For $p > 0.6$ the most infectious 20% of the runs are now resulting in $> 85\%$ of the total population becoming infected. This result shows us that, even for a simple network, small changes in the probability of transmission can have a large impact on the total number of infections.

Disordered geographic network model

Now that we have examined a branching process on simple version of the geographic network, we turn our attention to a more complex version of the same model. In this section we modify the 2D lattice network to allow connections to occur between nodes that are not nearest neighbours. This is an important factor to consider in the study of disease propagation as in social systems connections are generally not confined to nearest neighbours. To do this we assign each agent a finite probability of connecting to every other agent in the network. Where this probability depends on the distance d between each pair of nodes/agents. We will

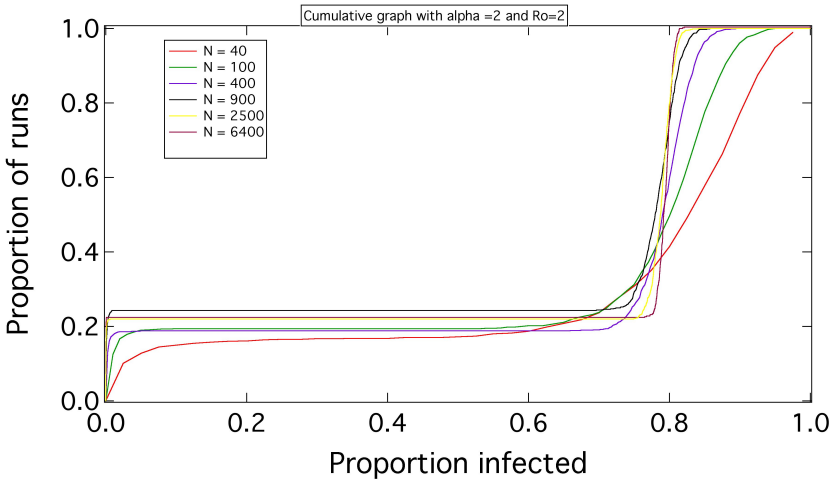

Figure 9: Cumulative distribution for long distance geographic network for various population sizes. Vertical axis shows the proportion of runs and horizontal axis displays n_β the proportion infected. Here $\alpha = 2$ and R_0 is set to be equal to 2 for all population sizes.

call this model the long distance (LD) network model. This modelling approach aims to create a network where each agent is more likely to come into contact with agents near them and less likely to come into contact with agents located a long distance away from them. The agents are initially positioned on a square grid and the distance is calculated from this.[1] We defined the probability of connection between two agents i and j to be given by the following equation;

$$Pij = c(d_{ij})^{-\alpha}. \tag{7}$$

Where d_{ij} is the line of sight distance between the two agents i and j, c is the probability of connection (equivalent to p as described in the 2D lattice model) and α is the exponent that defines the likelihood of a long distance connection. As α increases the probability of long distance connections decreases, whilst the probability of nearest neighbour connections increases. Indeed as α tends towards infinity we are able to get the 2D lattice network as a special realisation of this model. This long-distance model creates a network of connections amongst all the agents, and it is this network that we use to run the branching process on.

We first investigated the effect of population size on the size of the disease outbreak. To do this we varied the population size and for each population n we ran the simulation 100,000 times. We did this for $\alpha = 2$ so that there would be a significant number of long range connections, and $R_0 = 2$ for direct comparison to the 2D lattice network model. The cumulative distribution is plotted in Fig.9 for n between 40 and 6400. From this set of results we can see that there are two distinct types of disease outbreak. The first set (Type

[1]There are a number of different ways to calculate distance between agents depending on the system being modelled, here we chose straight line distance, though the approach could easily be extended to a grid based manhattan-type distance.

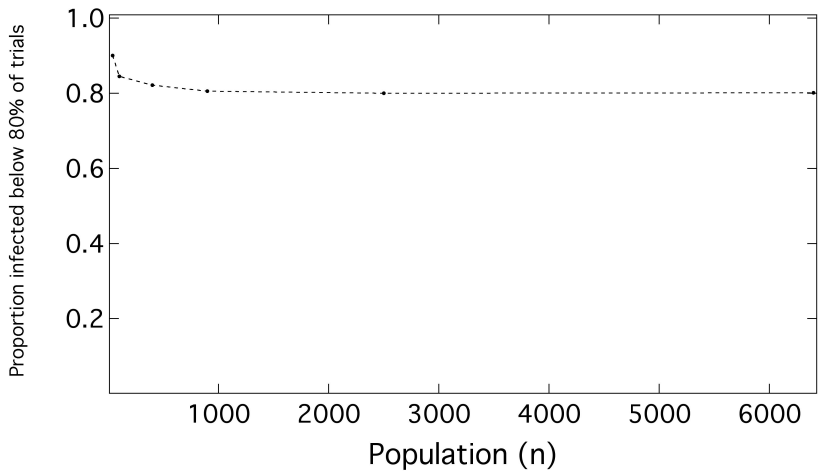

Figure 10: Graph showing the dependence of N^{80} on population size n. The circles represent the 80th percentile calculated from 100,000 runs of the model.

I) is a low intensity outbreak that only infects a very small number of individuals within the population (i.e. $N_\beta < 0.05$). The second type (Type II) of outbreak is a large scale outbreak that infects $\sim 80\%$ of the population. For large populations in the long distance network model these are the only two types of outbreaks, with no intermediate values of n_β present. From the cumulative distribution graph we can see that $\sim 20\%$ of the outbreaks are the low intensity type 1 outbreaks, whilst $\sim 80\%$ of the outbreaks are high intensity outbreaks. As n increases the cumulative distribution curve becomes more like a step function indicating again the presence of finite size effects within the network model.

We then calculated the 80^{th} percentile for each of the cumulative curves and plotted this on a separate graph for various values of n. These results are displayed in Fig.10 and from this figure we can now compare our results for the long distance network model with the 2D lattice model. In the long distance (LD) model we see that N^{80} does not depend strongly on the size of the population. With the top 20% of the runs infecting almost the same proportion of the population for $n = 100$ to $n = 6400$. This result is in stark contrast to the strong n dependence of N^{80} in the 2D network model (see Fig.7). This result means that for the LD model once a disease outbreak has moved above the threshold level it will infect a large proportion of the population regardless of the total size of the population.

In the next set of simulations we kept fixed the size of the population at $n = 6400$ and varied the probability of transmission (c). The results from these simulations are displayed in Fig.11 where the curve in the top left corner represent the data from the smallest value of c and the curve in the bottom right corner the largest value of c. The data is plotted on a cumulative distribution graph and we can again see the characteristic stepwise function present for all curves, although it is most apparent for the curves with high transmission probabilities (c¿0.1). The stepwise curve is indicative of the two types of disease outbreaks

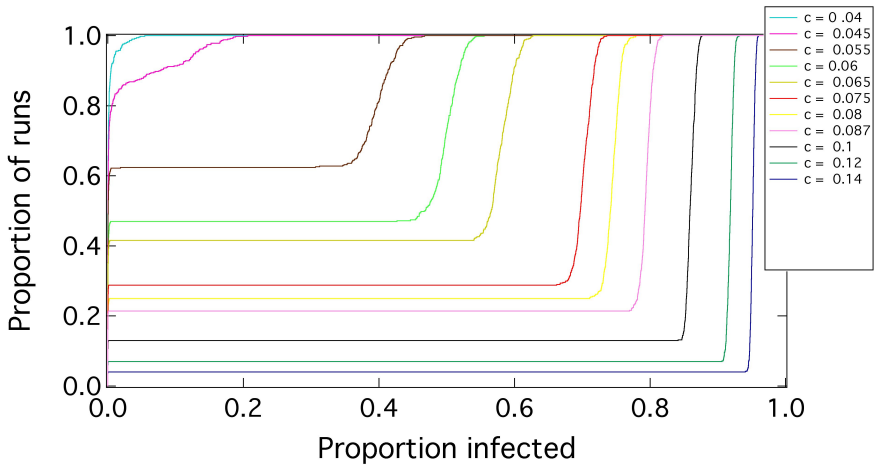

Figure 11: Graph showing the cumulative distribution for the LD model with $\alpha = 2$ and $n = 6400$ for all the curves. The probability of transmission is varied here from $c = 0.04$ which appears on the upper left side to $c = 0.14$ which is the curve on the lower right side. The horizontal axis represents the proportion infected (n_β) and the vertical axis represents the cumulative proportion of runs.

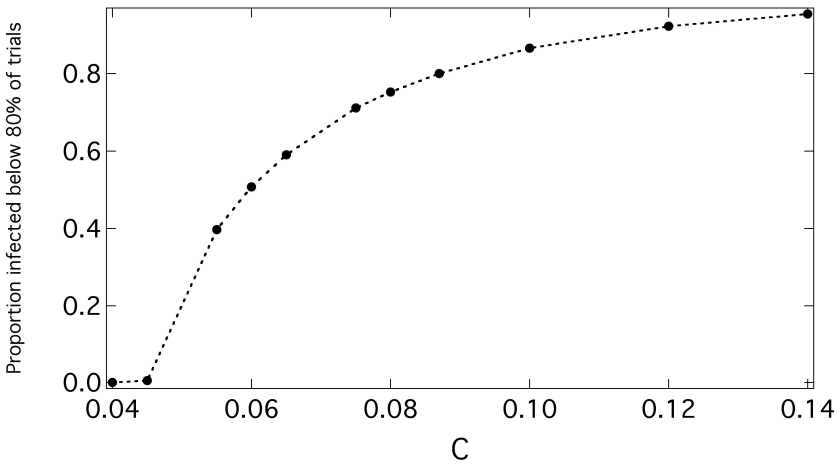

Figure 12: Graph showing the dependence of N^{80} on c the probability of disease transmission. The circles here represent the 80th percentile calculated from 100,000 iterations of the model with n fixed at 6400 nodes.

that we talked about above. Here we see that by increasing the size of c we can increase the proportion n_β of the population that is infected in the type II outbreak. Increasing c also increases the proportion of type II outbreaks that infect a large proportion of the population.

The 80th percentile was calculated for each of these graphs shown in Fig.11 and these results are displayed in Fig.12. We note that for low values of c, small increases in the transmission probability can make large differences in the total proportion of the population infected. This behaviour is similar in nature to that observed for increasing values of p in the 2D lattice model.

Conclusion

In combining a geographical and network structure to standard epidemiological models some interesting results have appeared when we adjusted the scaling parameter in the network we found that it doesnt take a large R_0 to infect a substantial proportion of a population due to the long range connections. We also found that when we varied the size of the population that this had little impact on the proportion of the population infected unlike standard models.

The combination of a geographical and network structure to standard epidemiological models could be extended to: include analytic results to accompany the numeric results; compare this work with empirical data comparison to empirical data; to look at clustering within the network and in particular to look at the possibility of assigning long distance connections to agents that already have a high number of connections.

References

[1] Watts D J and Strogatz S. H. Collective dynamics of 'small-world' networks, Nature, 393:440-442 (1998).

[2] Barabsi, A-L. and Albert, R. "Emergence of Scaling in Random Networks." Science 286, 509-512, 1999.

[3] Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. and Getz, W. M. Superspreading and the impact of individual variation on disease emergence. Nature 438, 2005, 355359.

[4] Newman, M. J., A. L. Barabasi, and D. J. Watts, editors. 2004. The structure and dynamics of complex networks. Princeton University Press, Princeton, New Jersey, USA.

[5] May R.M., Anderson R.M. 1987 Transmission dynamics of HIV infection. Nature 326, 137-42.

[6] James, A, Pitchford, J, Plank M.J. An event-based model of superspreading in epidemics, Proceedings of the Royal Society Series B 274 (2007) 741-747.

[7] Milgram, S., Small World Problem. Psychology today, May 1967. pp 60 67.